

Comment

Towards a theory of word order. Comment on “Dependency distance: a new perspective on syntactic patterns in natural language” by Haitao Liu et al

R. Ferrer-i-Cancho

*Complexity & Quantitative Linguistics Lab, LARCA Research Group, Departament de
Ciències de la Computació, Universitat Politècnica de Catalunya, Campus Nord, Edifici
Omega Jordi Girona Salgado 1-3. 08034 Barcelona, Catalonia (Spain)*

Liu et al’s reflections on the term dependency length minimization [1] may look anecdotal but they are not. By the turn of the 20th century, we put forward a “Euclidean distance minimization” hypothesis for the distance between syntactically linked words and various word order phenomena [2, 3]¹. Later on, pressure from language researchers forced us to replace it with terms such as “online memory minimization” [5] because our initial formulation was obscure to them. Recently, researchers from all over the world have been granted to use the term “dependency length minimization” by the popes thanks to whom [6] came into light. Although “length” is a particular case of distance in this context and thus downsizes our original formulation, it is still abstract enough to allow for progress in theoretical research [7] and frees us from the heavy burden of contingency, *i.e.* the real implementation of the principle (at present believed to result from decay and interference as reviewed by Liu et al) or the current view of the architecture of memory [8, 9]. Our position is grounded on the high predictive power of that principle *per se* [5].

However, the lower generality of the term “dependency length” can be an

Email address: rferrericanch@cs.upc.edu (R. Ferrer-i-Cancho)

¹These were pieces of our PhD thesis [4] that were submitted for publication before its defense.

obstacle to the construction of a fully-fledged scientific field [10]. First, distance minimization allows one to unify pressure to reduce dependency lengths (still distances) with constraints on word order variation and change arising from a principle of swap distance minimization [11]. “Distance minimization” has therefore a higher predictive power and greater utility in a general theory of communication. Second, distance provides a “formal background” or a “specific background” (following Bunge’s terminology [10]) from physics or mathematics such as the theory of geographical or spatial networks (where the syntactic dependency structures of sentences are particular cases in one dimension) [12, 13] or the theory for the distance between like elements in sequences (where the couple of words involved in a syntactic dependency are particular cases of like elements) [14]. Therefore we agree with [1] on the convenience of the term distance.

A less flashy contribution of [6] has been promoting the need of controlling for sentence length (as a predictor of dependency length in their mixed-effects regression model) in research on dependency length minimization, an important methodological issue [15] that was addressed early [2] but neglected in subsequent research (*e.g.*, [16, 17, 18]).

Liu et al focus their review on the fundamental principle of dependency length minimization but understanding how it interacts with other principles is vital. In 2009, we put forward another fundamental word order principle, *i.e.* predictability maximization, and presented a theoretical framework culminating in a conflict between dependency length minimization and predictability maximization [19]. For sociological reasons, these arguments started appearing in print many years later [20, 5, 21]. For the case of a single head and its dependents, the minimization of dependency lengths yields that the head should be placed at the center of the sequence whereas the principle of predictability maximization (or uncertainty minimization) yields that the head should be placed at one of the ends of the sequence (last if the head is the target of the prediction; first otherwise) [21, 20].

Liu et al review two major sources of evidence of dependency length minimization: the analysis of dependency treebanks and psychological experiments. A critical difference between them is that the former is based on the calculation of the total cost of the sentence (as a sum or mean of all the dependency lengths of the sentence) while the latter is based on a partial calculation and thus it can be misleading. Suppose that one wishes to compare the cost of two orderings of the same sentence. The observation that the processing cost of a sentence decreases when the length of a dependency

increases, does not allow one to conclude that dependency length minimization cannot explain the results because shortening an edge implies moving at least one of the words defining it, and every move could imply the reduction of other edges eventually reducing the total sum of dependency lengths or altering the so-called complexity profile (*e.g.*, [22]), rendering fair comparison impractical. The problem of partial calculation of length costs has already been discussed in the context of research on the cost of crossing dependencies [23] and worsens when the sentences being compared differ not only in order but also in content. Another challenge is the precision of dependency length that is typically measured in words. Lengths in phonemes or syllables shed light on why SVO languages show SOV order when the object is a short word such as a clitic [24].

Without addressing these issues, the anti-locality effects or long-distance dependencies reviewed by Liu et al can neither be attributed to predictability maximization nor be interpreted as a violation of dependency length minimization safely; an effective evaluation of the theoretical framework above can be impossible (as that framework makes theoretical predictions based on the calculation of full length costs). The real challenge for psycholinguistic research is not the extent to which the theoretical framework above is supported by current results in the lab but rather to increase the precision of dependency length measurements and investigate the experimental conditions in which the following theoretical predictions are observed[20, 21]: one principle beating the other, coexistence, collaboration between principles or the very same trade-off causing the delusion that word order constraints have relaxed dramatically or even disappeared. This is the way of physics.

Our concern for units of measurement is not a simple matter of precision but one of great theoretical importance: if the length of a dependency is measured in units of word length (*e.g.*, syllables or phonemes) then it follows that the length of a dependency will be strongly determined by the length of the words defining the dependency and that of the intermediate words. Therefore, pressure to reduce dependency lengths implies pressure for compression [25, 26], linking a principle of word order with a principle that operates (non-exclusively) on individual words. An understanding of how the principle of dependency length minimization interacts with other highly predictive principles beyond word order is a fundamental component of a general theory of animal behavior that has human language as a particular case.

Acknowledgments

We thank C. Gómez-Rodríguez and H. Liu for helpful comments. This research was funded by the grants 2014SGR 890 (MACDA) from AGAUR (Generalitat de Catalunya) and also the APCOM project (TIN2014-57226-P) from MINECO (Ministerio de Economía y Competitividad).

References

- [1] H. Liu, C. Xu, J. Liang, Dependency distance: a new perspective on syntactic patterns in natural languages, *Physics of Life Reviews* (2017) in this issue.
- [2] R. Ferrer-i-Cancho, Euclidean distance between syntactically linked words, *Physical Review E* 70 (2004) 056135.
- [3] R. Ferrer-i-Cancho, Some word order biases from limited brain resources. A mathematical approach, *Advances in Complex Systems* 11 (3) (2008) 393–414.
- [4] R. Ferrer-i-Cancho, Language: universals, principles and origins, Ph.D. thesis, Universitat Politècnica de Catalunya, Barcelona (2003).
- [5] R. Ferrer-i-Cancho, The placement of the head that minimizes online memory. A complex systems approach, *Language Dynamics and Change* 5 (2015) 114–137. doi:10.1163/22105832-00501007.
- [6] R. Futrell, K. Mahowald, E. Gibson, Large-scale evidence of dependency length minimization in 37 languages, *Proceedings of the National Academy of Sciences USA* 112 (33) (2015) 10336–10341.
- [7] R. Ferrer-i-Cancho, A stronger null hypothesis for crossing dependencies, *Europhysics Letters* 108 (2014) 58003.
- [8] J. Jonides, R. Lewis, D. E. Nee, C. Lustig, M. Berman, K. S. Moore, The mind and brain of short-term memory, *Annual Review of Psychology* 59 (2008) 193–224.
- [9] R. Lewis, S. Vasishth, J. Van Dyke, Computational principles of working memory in sentence comprehension, *Trends in Cognitive Sciences* 10 (2006) 447–454.

- [10] M. Bunge, What is pseudoscience?, *The Skeptical Inquirer* 9 (1984) 34–46.
- [11] R. Ferrer-i-Cancho, Kauffman’s adjacent possible in word order evolution, in: *The evolution of language: Proceedings of the 11th International Conference (EVLANG11)*, 2016.
- [12] R. Cohen, S. Havlin, *Complex Networks. Structure, Robustness and Function*, Cambridge University Press, Cambridge, UK, 2010, Ch. 8. Distances in geographically embedded networks, pp. 88–90.
- [13] M. T. Gastner, M. E. J. Newman, Optimal design of spatial distribution networks, *Phys. Rev. E* 74 (2006) 016117.
- [14] P. Zörnig, The distribution of the distance between like elements in a sequence, *Glottometrika* 6 (1984) 1–15.
- [15] R. Ferrer-i-Cancho, H. Liu, The risks of mixing dependency lengths from sequences of different length, *Glottology* 5 (2014) 143–155.
- [16] H. Liu, Dependency distance as a metric of language comprehension difficulty, *Journal of Cognitive Science* 9 (2008) 159–191.
- [17] D. Temperley, Dependency-length minimization in natural and artificial languages, *Journal of Quantitative Linguistics* 15 (3) (2008) 256–282.
- [18] D. Gildea, D. Temperley, Do grammars minimize dependency length?, *Cognitive Science* 34 (2) (2010) 286–310.
- [19] R. Ferrer-i-Cancho, Memory versus predictability in syntactic dependencies, Talk at the Kickoff Meeting Linguistic Networks (Bielefeld University, Germany) in June 5, 2009.
- [20] R. Ferrer-i-Cancho, Why might SOV be initially preferred and then lost or recovered? A theoretical framework, in: E. A. Cartmill, S. Roberts, H. Lyn, H. Cornish (Eds.), *THE EVOLUTION OF LANGUAGE - Proceedings of the 10th International Conference (EVLANG10)*, Wiley, Vienna, Austria, 2014, pp. 66–73, Evolution of Language Conference (Evolang 2014), April 14–17.

- [21] R. Ferrer-i-Cancho, The placement of the head that maximizes predictability. An information theoretic approach, <https://arxiv.org/abs/1705.09932>.
- [22] G. Morrill, *Categorial grammar: logical syntax, semantics, and processing*, Oxford University Press, Oxford, 2010.
- [23] R. Ferrer-i-Cancho, Non-crossing dependencies: least effort, not grammar, in: A. Mehler, A. Lücking, S. Banisch, P. Blanchard, B. Job (Eds.), *Towards a theoretical framework for analyzing complex linguistic networks*, Springer, Berlin, 2016, pp. 203–234.
- [24] R. Ferrer-i-Cancho, Reply to the commentary "Be careful when assuming the obvious", by P. Alday, *Language Dynamics and Change* 5 (2015) 147–155. doi:10.1163/22105832-00501009.
- [25] R. Ferrer-i-Cancho, A. Hernández-Fernández, D. Lusseau, G. Agaramoorthy, M. J. Hsu, S. Semple, Compression as a universal principle of animal behavior, *Cognitive Science* 37 (8) (2013) 1565–1578.
- [26] M. L. Gustison, S. Semple, R. Ferrer-i-Cancho, T. Bergman, Gelada vocal sequences follow Menzerath’s linguistic law, *Proceedings of the National Academy of Sciences USA* 13 (2016) E2750–E2758. doi:10.1073/pnas.1522072113.